中国人民大學
RENMIN UNIVERSITY OF CHINA

高瓴人工智能学院
Gaoling School of Artificial Intelligence

Inverse Game Theory for Stackelberg Games

The Blessing of Bounded Rationality

Weiran Shen
Renmin University of China

Joint work with:

Jibang Wu
University of Chicago

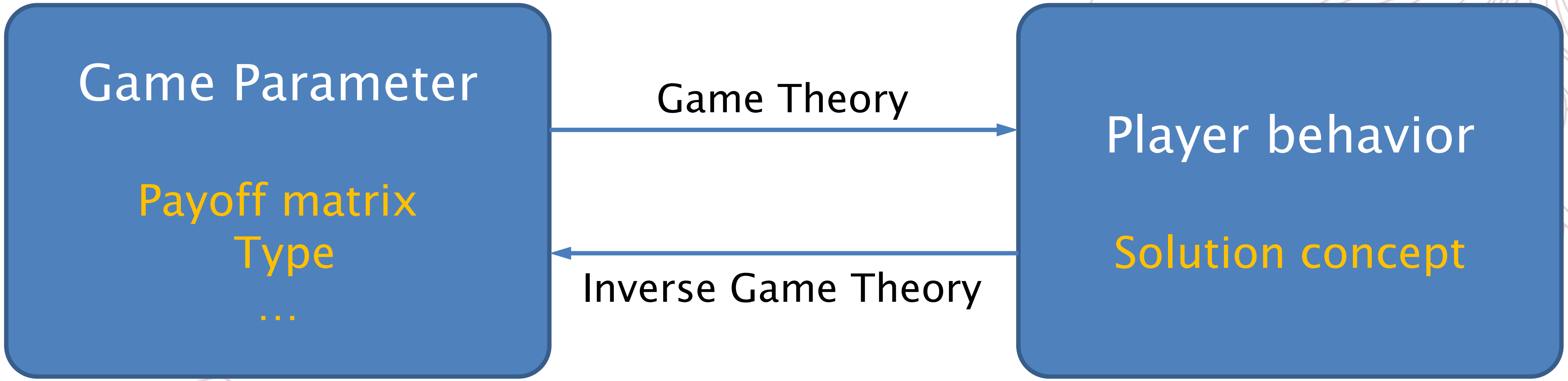
Fei Fang
Carnegie Mellon University

Haifeng Xu
University of Chicago

Motivation

- Game theory
  - Given game setting, predict players' behaviors
- In reality
  - E-commerce platform
    - Does not know customers' preferences, only observes their behaviors
  - Security domain
    - Does not know attackers' utility, only observes their responses

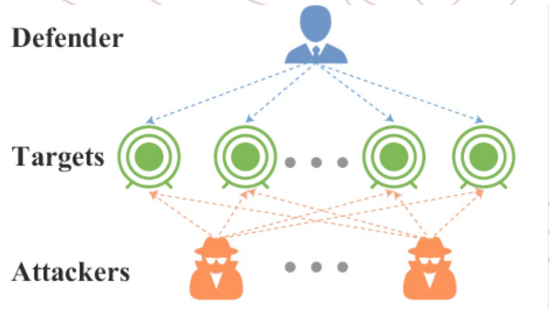
Inverse Game Theory



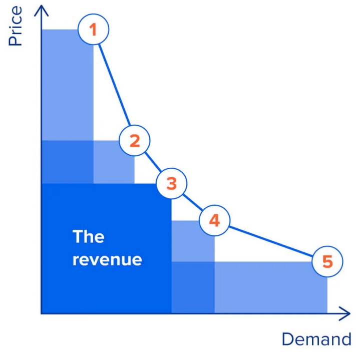
- Given equilibrium behaviors, what game parameters can induce such behaviors?

Setting

- Inverse Stackelberg game
  - A leader: commits to a strategy
  - A follower: responds to leader
  - Notations:
    - $U \in \mathbb{R}^{m \times n}$: leader's payoff
    - $V \in \mathbb{R}^{m \times n}$: follower's payoff
    - $x \in \Delta_m$: leader's strategy
    - $y \in \Delta_n$: follower's strategy



Optimal defense against strategic attackers



Optimal pricing against strategic buyers

Setting

- Inverse Stackelberg game
  - Leader can choose any mixed strategy x
  - Follower uses quantal response
    - Probability of choosing action j: $y_j = \frac{\exp(\lambda x^T V_j)}{\sum_{k\in[n]} \exp(\lambda x^T V_k)}$
    - Capture the follower's bounded rationality
  - Can the leader recover V by "querying" follower's response with x?

Quantal Response vs Best Response

- Best response
  - Computing the optimal leader strategy is simple
  - Recovering follower payoff is difficult
- Quantal response
  - Computing the optimal leader strategy is difficult
  - Recovering follower payoff is easy

Identifiability Issue

- Quantal response

$$y_j = \frac{\exp\left(\lambda x^T V_j\right)}{\sum_{j' \in [n]} \exp\left(\lambda x^T V_{j'}\right)} = \frac{\exp\left(\lambda \sum_k x_k V_{kj}\right)}{\sum_{j' \in [n]} \exp\left(\lambda \sum_k x_k V_{kj'}\right)}$$

  - y_i stays the same if we replace V_{kj} with $V_{kj} + c_k$
  - Row-wise translation leads to the same behavior!

- Logit distance

$$\Phi(V, \tilde{V}) = \frac{1}{mn} \sum_{i \in [m]} \min_z \left\| V_i - \tilde{V}_i - z \right\|_1$$

  - $\Phi(V, \tilde{V}) = 0$: perfect recovery of V

Learning From Mixed Strategies

- Every query x returns a mixed strategy y

Proposition (m strategies to success)

V can be perfectly recovered with m linearly independent queries.

- For any $\tilde{V}$, we can predict the response $\tilde{y}$ of x
  - Find a $\tilde{V}$ to match $\tilde{y}$ and y

Learning From Mixed Strategies

- Minimize the cross entropy between $\tilde{y}$ and y

$$\min -\sum_t \left[\sum_j y_j(t) \log \frac{\exp\left(\lambda x^T(t)\tilde{V}_j\right)}{\sum_{j'} \exp\left(\lambda x^T(t)\tilde{V}_{j'}\right)}\right]$$

⇓

$$min \quad \sum_t \left[\log\left(\sum_j \exp\left(z_j(t)\right)\right) - y(t)z(t)\right]$$

$$s.t. \quad z(t) = \lambda x^T(t)\tilde{V}$$

Convex!

Learning From Realized Actions

- Every query x returns an action y sampled from the quantal response model
- First thought
  - MLE: given queries $x(t)$, what $\tilde{V}$ leads to highest probability of observing $y(t)$?
    - Difficult to optimize
    - Difficult to bound error

Learning From Realized Actions

- Idea
  - Mixed strategy estimation: query x multiple times
  - Payoff estimation: use estimated response $\hat{y}$ to recover $\tilde{V}$
- Error bound
  - Bound the error of $\hat{y}$ with the number of queries
  - Bound the recovered $\tilde{V}$ given the error of $\hat{y}$

Learning From Realized Actions

- Mixed strategy estimation error

Lemma

For any query x, Let y be the underlying quantal response. Denote by $\rho = \min_i y_i$. With $O\left(\frac{\log(n/\delta)}{\rho\epsilon^2}\right)$ repeated queries of x, the empirical distribution $\hat{y}$ is a $(1-\epsilon)$-approximation of y with probability at least $1-\delta$.

Learning From Realized Actions

- Proof
  - Let $X_k = I(\text{response of query } k \text{ is action } i), \forall 1 \leq k \leq \frac{3\log(2n/\delta)}{y_i\epsilon^2}$
  - Let $X = \sum_{k\in[N]} X_k$. Then $\mu = \mathrm{E}[X] = \frac{3\log(2n/\delta)}{y_i\epsilon^2} y_i = \frac{3\log(2n/\delta)}{\epsilon^2}$
  - Chernoff multiplicative bound:

$$\Pr\{|X-\mu| > \epsilon\mu\} \leq 2\exp\left(-\epsilon^2 \frac{3\log\left(\frac{2n}{\delta}\right)}{3\epsilon^2}\right) = \frac{\delta}{n}$$

Relative error larger than ϵ

Learning From Realized Actions

- Proof
  - Using union bound, with probability at least $1-\delta$, $\frac{\hat{y}_i}{y_i} \in [1-\epsilon, 1+\epsilon] \subset \left[1-\epsilon, \frac{1}{1-\epsilon}\right], \forall i \in [n]$

Learning From Realized Actions

- Payoff recovery error

Lemma

There exists an algorithm that can recover V within the logit distance $\Phi(V, \tilde{V}) = O(\epsilon/\lambda)$ from m queries of $(1-\epsilon)$-multiplicative approximation of the follower's mixed strategies.

Learning From Realized Actions

- Proof
  - Let $\tilde{y}(t)$ be the estimated mixed strategy of the t-th query
  - Let $\beta_{jt} = \frac{\tilde{y}_j(t)}{y_j(t)} \in \left[1-\epsilon, \frac{1}{1-\epsilon}\right]$
  - Still solve:

$$
\begin{aligned}
min \quad & \sum_t \left[\log\left(\sum_j \exp\left(z_j(t)\right)\right) - \tilde{y}(t) z(t)\right] \\
s.t. \quad & z(t) = \lambda x^T(t) \tilde{V}
\end{aligned}
$$

Learning From Realized Actions

- Proof
  - Solution satisfies:

$$\tilde{V} = V + \frac{1}{\lambda}(X^{-1})^T \log \beta + c$$

$X = [x(t)]_{t\in[m]}$
Full rank matrix

Element-wise log

Row-wise translation

Learning From Realized Actions

- Proof
  - Solution satisfies:

$$\Phi(V,\tilde{V}) = \frac{1}{mn}\sum_{i\in[m]} \min_z \left\|V_i - \tilde{V}_i - z\right\|_1$$

$$\le \frac{1}{mn}\left\|\frac{1}{\lambda}(X^{-1})^T \log\beta\right\|_1$$

$$= \frac{1}{mn}\left\|\frac{1}{\lambda}(X^{-1})^T\right\|_1 mnO(\epsilon)$$

$$= O\left(\frac{\epsilon}{\lambda}\right)$$

$$\beta_{jt} \in \left[1-\epsilon, \frac{1}{1-\epsilon}\right]$$

Choose X to be the identity matrix

Learning From Realized Actions

- Leader utility bound

Theorem (informal)

Under certain technical conditions, we can construct an nearly optimal leader strategy for any $\tilde{V}$ with $\Phi(V, \tilde{V}) = O(\epsilon/mn)$

Summary & Future Work

- Summary
  - Inverse Stackelberg game
  - V can be recovered using m follower mixed strategies
  - Sample complexity of learning V
- Future work
  - More general settings
  - Other bounded rationality model
  - Choose queries in a smarter way

Thanks!

Q & A

Weiran Shen
Renmin University of China

shenweiran@ruc.edu.cn